

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:08:59 ; Search time 299.73 Seconds
(without alignments)
16.815 Million cell updates/sec

Title: US-09-331-631a-1_COPY_74_116
Perfect score: 248
Sequence: 1 NOEDPOTECQCCORRCROE.....ROOYCORCKEICEEEY 43

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMEL_15:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248	100.0	666	10 Q9SP15	Q9SP15 macadamia i
2	242	97.6	666	10 Q9SP14	Q9SP14 macadamia i
3	241	97.2	625	10 Q9SP13	Q9SP13 macadamia i
4	119	48.0	593	10 Q9SEW4	Q9SEW4 juglans reg
5	117	47.2	525	10 Q43358	Q43358 theobroma c
6	96	38.7	411	5 P91419	P91419 caenorhabdi
7	80.5	32.5	425	5 Q17400	Q17400 caenorhabdi
8	80.5	32.5	600	5 Q17401	Q17401 caenorhabdi
9	80.5	32.5	810	10 Q92W13	Q92W13 cucurbita m
10	79	31.9	572	5 Q19594	Q19594 caenorhabdi
11	77.5	31.2	242	5 Q19919	Q19919 caenorhabdi
12	75	30.2	339	5 Q9N4R1	Q9N4R1 caenorhabdi
13	74	29.8	388	5 Q16500	Q16500 caenorhabdi
14	74	29.8	388	5 Q16501	Q16501 caenorhabdi
15	74	29.8	438	5 Q16502	Q16502 caenorhabdi
16	74	29.8	445	5 Q16511	Q16511 caenorhabdi
17	73.5	29.6	335	5 Q9NF70	Q9NF70 caenorhabdi
18	73.5	29.6	335	5 Q9NF69	Q9NF69 caenorhabdi
19	73.5	29.6	419	5 Q9NA38	Q9NA38 caenorhabdi

20	73.5	29.6	420	5 Q9NA61	Q9NA61 caenorhabdi
21	73	29.4	388	5 Q44606	Q44606 caenorhabdi
22	72.5	29.2	330	5 Q18118	Q18118 caenorhabdi
23	72.5	29.2	709	5 Q9NA42	Q9NA42 caenorhabdi
24	71.5	28.8	335	5 Q9NA63	Q9NA63 caenorhabdi
25	70.5	28.4	378	5 Q27383	Q27383 caenorhabdi
26	70	28.2	1513	5 Q17970	Q17970 caenorhabdi
27	68.5	27.6	393	10 Q92TP0	Q92TP0 oryza sativ
28	68.5	27.6	654	5 Q17982	Q17982 caenorhabdi
29	68	27.4	594	5 Q77337	Q77337 plasmodium
30	66.5	26.8	273	5 Q45362	Q45362 caenorhabdi
31	66.5	26.8	314	5 Q23390	Q23390 caenorhabdi
32	66.5	26.8	356	5 Q17316	Q17316 ceratilis c
33	66.5	26.8	637	10 Q03678	Q03678 hordeum vul
34	65	26.2	111	5 Q19054	Q19054 caenorhabdi
35	65	26.2	539	4 Q9NUA2	Q9NUA2 homo sapien
36	65	26.2	1306	5 Q77273	Q77273 drosophila
37	63	25.4	243	5 Q9NL90	Q9NL90 entamoeba d
38	63	25.4	910	11 Q54899	Q54899 mus musculu
39	63	25.4	910	11 Q88704	Q88704 mus musculu
40	62.5	25.2	342	4 Q9UK28	Q9UK28 homo sapien
41	62	25.0	425	5 Q15755	Q15755 dictyosteli
42	62	25.0	648	5 Q9TXB8	Q9TXB8 dictyosteli
43	62	25.0	715	4 Q9UEX4	Q9UEX4 homo sapien
44	62	25.0	1080	5 Q9NZM8	Q9NZM8 drosophila
45	61	24.6	400	6 Q95337	Q95337 tupia glis

ALIGNMENTS

RESULT 1
Q9SP15 PRELIMINARY; PRT; 666 AA.

AC Q9SP15; 01-MAY-2000 (Tremblrel, 13, Created)
DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel, 15, Last annotation update)
DE VICILIN PRECURSOR.
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OC NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUIT KERNEL;
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
globulin protein in Macadamia integrifolia kernels.";
RL Plant J. 0:0-0(1999).
DR EMBL: AF161883; AAD54244.1; -
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR00113; -
DR PFWM: PF00546; SeedStore-7s; 1.
SQ SEQUENCE 666 AA; 78217 MW; C752B884B2DF0224 CRC64;

Query Match 100.0%; Score 248; DB 10; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NOEDPOTECQCCORRCROESGPRROOYCORCKEICEEEY 43
DB 74 NOEDPOTECQCCORRCROESGPRROOYCORCKEICEEEY 116

RESULT 2
Q9SP14 PRELIMINARY; PRT; 666 AA.
AC Q9SP14; 01-MAY-2000 (Tremblrel, 13, Created)
DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)

DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR.
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta;
OC Magnoliophyta: eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUIT KERNEL;
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia."
RL Plant J. 0:0-0(1999).
DR EMBL: AF161884; AAD54245.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
SQ SEQUENCE 666 AA; 78243 MW; 0ECA22F8710F8A7B CRC64;

Query Match 97.6%; Score 242; DB 10; Length 666;
Best Local Similarity 95.3%; Pred. No. 6.5e-22;
Matches 41; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 NOEDPOTECOCORRCRQESGPRQOQYCORRKEICEEEY 43
DB 74 NODDPOTDCOCORRCRQESGPRQOQYCORRKEICEEEY 116

RESULT 3
O9SPL3 PRELIMINARY; PRT; 625 AA.
ID O9SPL3;
AC O9SPL3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR (FRAGMENT).
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta;
OC Magnoliophyta: eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUIT KERNEL;
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia."
RL Plant J. 0:0-0(1999).
DR EMBL: AF161885; AAD54246.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
FT NON_TER 1
SQ SEQUENCE 625 AA; 73586 MW; 415808A89D370296 CRC64;

Query Match 97.2%; Score 241; DB 10; Length 625;
Best Local Similarity 97.7%; Pred. No. 8.2e-22;
Matches 42; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 NOEDPOTECOCORRCRQESGPRQOQYCORRKEICEEEY 43
DB 33 NODDPOTECOCORRCRQESGPRQOQYCORRKEICEEEY 75

RESULT 4
O9SEW4 PRELIMINARY; PRT; 593 AA.
ID O9SEW4;
AC O9SEW4;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE VICILIN-LIKE PROTEIN PRECURSOR (FRAGMENT).
OS Juglans regia (English walnut).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta;
OC Magnoliophyta: eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Fagales; Juglandaceae; Juglans.
OX NCBI_TaxID=51240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNLAND; TISSUE=SOMATIC EMBRYO LINE;
RA Teuber S.S., Jarvis K.C., Peterson W.R., Dandekar A.M., Ansari A.A.;
RT "Identification and cloning of a cDNA encoding a vicilin-like protein,
RT Jug r 2, from English walnut kernel (Juglans regia): a major food
RT allergen."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF066055; AAF18269.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
FT NON_TER 1
SQ SEQUENCE 593 AA; 69990 MW; 9BA127E19B18C0D8 CRC64;

Query Match 48.0%; Score 119; DB 10; Length 593;
Best Local Similarity 51.2%; Pred. No. 5.3e-07;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;
OY 1 NOEDPOTECOCORRCRQESGPRQOQYCORRKEICEEE 41
DB 10 NRPDPREQYRQCOYCRKQGGUQKQOQOQIRCEERLEED 50

RESULT 5
O43358 PRELIMINARY; PRT; 525 AA.
ID O43358;
AC O43358;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR.
GN CSV.
OS Theobroma cacao (Cacao).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta;
OC Magnoliophyta: eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Theobroma.
OX NCBI_TaxID=3641;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAVES;
RA McHenry L., Fritz P.J.;
RT MEDLINE=92288309; PubMed=1600151;
RT "Comparison of the structure and nucleotide sequences of vicilin genes
RT of cocoa and cotton raise questions about vicilin evolution."
RL Plant Mol. Biol. 18:1173-1176(1992).
DR EMBL: X62625; CAA44493.1; -.
DR EMBL: X62626; CAA44494.1; -.
DR HSSP: P02853; 2PHL.
DR MENDEL: 30919; Thecc:1188; 30919.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
DR PRODOM: PD081059; -.
FT SIGNAL 1
FT CHAIN 25 525 POTENTIAL.
FT CHAIN 25 525 VICILIN.
SQ SEQUENCE 525 AA; 60798 MW; 19114C05C248905D CRC64;

Query Match 47.2%; Score 117; DB 10; Length 525;
Best Local Similarity 50.0%; Pred. No. 8.4e-07;
Matches 20; Conservative 10; Mismatches 10; Indels 0; Gaps 0;
OY 2 QEDPOTECOCORRCRQESGPRQOQYCORRKEICEEE 41
DB 1 QEDPOTECOCORRCRQESGPRQOQYCORRKEICEEE 41

DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PVI000.
OS *Cucurbita maxima* (Pumpkin) (Winter squash).
OC Euarystola: Viridiplantae: Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta: euclidocyledons: core eudicots; Rosidae; eurosids I;
OC Cucurbitales: Cucurbitaceae; Cucurbita.
OX NCBI_TaxID=3661;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=KUBOKAWA AMAKURI NANKIN: TISSUE=COTYLEDON;
RX MEDLINE=99107919: Pubmed=9891029;
RA Yamada K., Shimada T., Kondo M., Nishimura M., Hara-Nishimura I.;
RT "Multiple functional proteins are produced by cleaving Asn-Gln bonds
of a single precursor for vacuolar processing enzyme.";
RL J. Biol. Chem. 274:2563-2570(1999).
DR EMBL: AB019195; BAA34056.1; "
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; "
DR Pfam: PF00546; Seedstore_7s; 1.
DR PRODOM: PD081059; "
SO SEQUENCE 810 AA; 97314 MW; A829A3F7542266AB CRC64;

```

QY      1 NOE-DPOTCCQQCRRCROEGSPGQQOYCQRCEICEEEFE 42
        |||::||::||::||::||::||::||::||::||::||
Db      69 NORGPSRAIEYECRLRICOVAERGVDEQRCDEVCERLERERO 111
          Matches 16; Conservative 9; Mismatches 17; Indels 1; Gaps 1
Query Match 32.5%; Score 80.5; DB 10; Length 810;
Best Local Similarity 37.2%; Pred. No. 0.032;

```

Query Match	Best Local Similarity	Score	DB	Length
Matches 17; Conservative	31.9%;	79;	DB 5;	572;
	Pred. No. 0.037;			
	8; Mismatches	7;	Indels	18; Gaps
				3

QY	8	EQQCCRCRCRQ-----	QESGPRDQYQCRCKE	-ICDEE	40
		:: :			
Db	381	QCCQCCNTCCQCFAPVCEQCCASICHCQPSAPCCQ	-CQNTCCQCFAPVCEQ		429
		:: :			

RESULT 11	
Q19919	
ID Q19919	PRELIMINARY; PRT; 242 AA.

AC 019919; 01-JUN-1998 (TREMBLrel.06, Created)
DT 01-JUN-1998 (TREMBLrel.06, last sequence update)
DT 01-JUN-1998 (TREMBLrel.06, last annotation update)
DE 01-JUN-1998 (TREMBLrel.06, last annotation update)
DE HYPOTHETICAL 26.2 KDA PROTEIN F31A3.1 IN CHROMOSOME X.
GN F31A3.1.
GN Caenorhabditis elegans.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditioidea;
OC Rhabditidae; Pelodermaeae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;
RA Murray J., De T.T.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBD databases.
CC -1- SIMILARITY: BELONGS TO FAMILY UPF.
DR EMBL; U58742; AAB36856.1; -
DR WORMPEP; F31A3.1; CE07158.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 3
FT POTENTIAL 23
SQ SEQUENCE 242 AA; 26213 MW; C7D6596B00829317 CRC64;

Query Match	31.2%	Score 77.5	DB 5	length 242
Best Local Similarity	33.3%	Pred. No. 0.028		
Matches 17; Conservative	12	Mismatches	11	Gaps 4

RESULT	12			
Q9N4R1		PRELIMINARY:	PRM:	339 AA.
ID	Q9N4R1			
AC	Q9N4R1:			
DT	01-OCT-2000 (TREMBLrel, 15, Created)			
DT	01-OCT-2000 (TREMBLrel, 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel, 15, Last annotation update)			
DE	HYPOTHETICAL PROTEIN Y5H2A.C.			
GN	Y5H2A.C.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;			
CC	Rhabditidae; Peloiderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RX	MEDLINE=99069613; Pubmed=9851916;			
RA	None;			
RT	"Genome sequence of the nematode <i>C. elegans</i> a platform for			
RT	investigating biology. The <i>C. elegans</i> Sequencing Consortium.";			
RL	Science 282:2012-2018 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RA	Waterston R.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL: AC006809; AAF59626.1; -			
Q0	SEQUENCE 339 AA; 36218 MW; 40233423237C432D CRC64;			

Query Match 30.2%; Score 75; DB 5; Length 339;
Best Local Similarity 29.1%; Pred. No. 0.075;
Matches 16; Conservative 11; Mismatches 12; Indels 16; Gaps 3;

RESULT 13
016500

Fri Mar 2 09:28:55 2001

us-09-331-631a-1_copy_74_116.rspt

Page 6

		29.8%	Score	74;	D _B	5;	Length	438;	
Query Match									
Best Local Similarity		30.2%	Pred	No.	0.1%;				
Matches	13;	Conservative	12;	Mismatches	12;	Indels	6;	Gaps	2;
OY	3	EDPTECO-CCORRC-----ROESGPRQOQYCORKEICEE	39						
		: :							
D _B	191	QQAADACCPGCCQQQCOSCVQDQDPSTICCFACNMTCSDITCQ	233						

Search completed: March 1, 2001, 16:09:02
Job time: 1546 sec

-1_copy-74_116.rsp